

Occasio Digital Social History Archive

The Internet archive of the IISH
Jenneke Quast

Campaigning on the Internet

In 1994 the International Institute of Social History (IISH) launched a special project to safeguard a relatively new category of electronic documents that without immediate action was in danger of being lost forever. These were Internet messages from information networks dealing with social, political and ecological issues.

The Internet was designed at the height of the Cold War by and for the U.S. Department of Defence, which wanted its communications network to remain operational after a nuclear attack and sought to link all its research institutions. According to the terms of that era, the Internet is a 'rightist' invention. The left, however, was relatively quick to recognize the value of the Internet for disseminating information. By the mid 1980s many progressive groups used computer networks – often known as bulletin board systems - to distribute information about social and political issues. The number of action committees, solidarity movements and non-governmental organizations using this medium has increased considerably since then. The average producers are obviously more concerned with the immediate political objectives they write about than with long-term storage of their documents. Still, these documents are extremely vulnerable. All archives are at risk of being neglected or discarded entirely or partially over time. Organizations may also simply cease to exist, and their records disappear with them. Computer files lead a more tenuous existence even than paper ones. All too often, people learn after their computer has crashed that they have no back-up copies, or that the ones they do have are poor or incomplete. Even decent back-up files quickly become impossible to read because of obsolete hardware or software.

Since its establishment in 1935, the IISH (www.iisg.nl) has collected social- historical documents made of all conceivable types of material. Some texts circulating on the Internet would certainly be included in the collection as hard copies but are simply not available on paper. Examples of such documents include the news messages distributed by the Association for Progressive Communications (APC), a global electronic information network serving non-governmental organizations (NGOs) all over the world. Its member networks exchange e-mail and electronic *conferences* (i.e. newsgroups, similar to Usenet groups) with dozens of partner networks worldwide. The APC *conferences* concern issues such as human rights, labour relations, war and peace, development, as well as many other social, political and ecological topics and match the IISH collection profile perfectly. The message contents are highly relevant and virtually free of advertising. The APC is particularly interesting because of its use during the early years of the Internet by many international NGOs as an exclusive platform for their private and public newsletters and bulletins (www.apc.org).

The IISH decided to collect the APC newsgroup messages to safeguard them and make them available to researchers. The Institute works closely on this project with the Antenna Foundation, the Dutch representative of the APC for many years and one of the oldest Internet providers of the Netherlands. The project was allocated funding by the

Netherlands Organization for Scientific Research. It was named Occasio after an idea by Tjebbe van Tijen, who also first suggested the archives project (www.iisg.nl/~tvt/). In 1995 Antenna started to collect and store newsgroup messages for the IISH. They included materials from the second half of the 1980s, when information from NGOs on the Internet was still rare. As a result, 2,300,000 messages from 2,997 newsgroups dating from 1986-2002 are now accessible via the Occasio Web site (www.iisg.nl/occasio/).

A special collection of messages concerns the wars in the former Yugoslavia. When direct telephone communications between Croatia and Serbia became almost impossible, the American peace activist Eric Bachman proposed establishing e-mail contact between Zagreb and Belgrade using non-commercial computer networks in Germany. In 1992 the ZaMir ('For Peace') information network was established with support from German non-profit agencies and Antenna, which arranged training for the ZaMir staff. In 1998 ZaMir ceased to exist as a network. To safeguard their files, Eric Bachman visited all the former network *nodes* to make back-up copies of their hard drives. The ZaMir materials can now be accessed separately.

The archive

The entire collection is stored on hard drives. Back-up copies are made to avert loss of data, in case of a disk failure. Different physical carriers are used, as each is vulnerable in its own way. Hard drives crash and are vulnerable to hackers. Tapes (even expensive ones) can suddenly pulverize. Data storage on CD-ROMs is not reliable, because the writing process is not: one bad bit makes the entire disk unreadable. Moreover, CDs are subject to chemical erosion, unless they are treated with an expensive protective coating. That is why copies need to be made periodically on the different storage media, which are regularly replaced (tapes are renewed about once every three months). As an added precaution, copies of the entire Occasio collection are stored on computers in different parts of the world.

The numbers of documents involved exceed manual processing capacities. From the outset, therefore, collection and storage processes have been fully automated. News messages arrive automatically on a dedicated server, where they are stored in their original format to retain all information about each message, such as origin, name of the original sender, subject, date and time and a unique document code known as the message ID. They are made accessible in a database, where they can be searched according to words from the text, newsgroup names and keywords that the founders assigned to the groups at the time of establishment. No additional subject indexes are compiled. The messages were originally produced in a relatively simple electronic text format without graphics. Antenna has made them accessible with software that allows them to be viewed via a Web interface without changing the original material. Throughout the process, they used open-source software wherever possible.

Access and copyright

The archive is available online. Private APC conferences may be consulted only at the premises of the IISH, in keeping with the Institute's regulations. Although locating and securing the documents was the top priority, the partners in the project intended from the outset to make them accessible and available to the public on the basic assumption that documents published on the Internet are universally accessible, and that APC members would waive all copyrights for the sake of disseminating the information. Since copyright obviously still applies, Occasio Website visitors may use the documents in the Occasio archive exclusively for non-commercial individual research. Use for all other purposes requires contacting the author of the document concerned, as is the procedure for archive users at the IISH reading room. To accommodate potential objections, authors are entitled to block Web use of their messages. The documents will continue to be stored at the IISH and may be consulted there.

Future changes

As Occasio is an archival collection, users cannot reply to the messages or post new ones. Antenna and the IISH, however, will continue to collect new messages from newsgroups and mailing lists, which in due time will be added to the main archive. In the meantime, information technology will evolve, and the project partners look forward to using future hardware and software improvements to safeguard the archive and ensure continuous access in the coming years.

Since the start of the project, the Internet, especially the Web, has grown explosively, as has the Web presence of NGOs. While Antenna and the IISH considered several options for storing relevant Web pages, the level of completeness pursued with the APC is unlikely for the Web, if only because of the vast changes in progress elsewhere in this field. The International Internet Preservation Consortium, comprising Brewster Kahle's Internet Archive, the Library of Congress, the British Library and the national libraries of Australia and several European countries, is a very valuable initiative for preserving Internet content for future generations (netpreserve.org/about/index.php). And while Occasio is thus not the only Internet archive, it is unique by virtue of its international collection in a specific field covering an extended period.

February 2005